R E P O R T R E S U M E S

ED 003 591 56

SOUND FILM RECORDINGS IN IMPROVING CLASSROOM COMMUNICATIONS.

BY- MACCOBY, NATHAN AND OTHERS
STANFORD UNIV., CALIF., INST.FOR COMMUNIC. RES.

REPORT NUMBER NDEA-VIIA-680 PUB DATE 63
GRANT OEG-7-14-1380-136
EDRS PRICE MF-$0.18 HC-$3.28 82P.

DESCRIPTORS- *FILMS, SOUND FILMS, *TEACHER EVALUATION, TEACHER RATING, *TEACHER RESPONSE, *COMMUNICATION SKILLS, *COMMUNICATION PROBLEMS, VERBAL COMMUNICATION, VERBAL TESTS, VISUAL DISCRIMINATION, STANFORD, CALIFORNIA

THIS STUDY INVESTIGATED THE EXTENT TO WHICH SOUND FILM RECORDINGS OF PUPILS IN TEACHING-LEARNING SITUATIONS CAN BE USED AS A BASIS FOR TRAINING TEACHERS TO IMPROVE THEIR ACCURACY IN INTERPRETING NONVERBAL COMMUNICATIONS, AND THE EXTENT OF DIFFERENCES AMONG TEACHERS, NEW TEACHERS, AND NONTEACHERS IN THEIR ABILITY TO INTERPRET NONVERBAL COMMUNICATIONS. EQUIPMENT AND METHODS WERE DEVISED TO RECORD INSTANCES OF STUDENT CLASSROOM RESPONSE AND SEVERAL EXPERIMENTS CONDUCTED. SUBJECTS WERE SHOWN SOUND FILM CLIPS OF TEACHER-STUDENT INTERACTIONS AND WERE ASKED TO RATE THE AMOUNT OF STUDENT COMPREHENSION IN THE CLIPS. SOME SUBJECTS WERE SHOWN ONLY THE PICTURE TRACK, SOME HEARD ONLY THE SOUND, AND OTHERS SAW AND HEARD THE COMPLETE RECORDING. THE RESULTS INDICATE THAT VERBAL CUES CONTRIBUTE SIGNIFICANTLY TO THE ACCURACY OF JUDGMENT, WHILE NONVERBAL CUES, SPECIFICALLY VISUAL ONES, DO NOT INCREASE ACCURACY OF JUDGMENT. THERE WAS NO EVIDENCE THAT THE USUAL TEACHER TRAINING PROGRAM IMPROVED ACCURACY OF JUDGMENTS. FURTHER RESEARCH IN THE AREA WAS RECOMMENDED. (JC)

ED003591

Sound Film Recordings in Improving Classroom Communications

680

Experimental Studies in Non-verbal Communication

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Nathan Maccoby
Jon Jecker
Henry Breitrose
Ernest Rose

Title VII Project 680
National Defense Education Act of 1958
Grant 7-14-1380-136

Institute for Communication Research
Stanford University, Stanford, California

The Research Reported Herein Was Supported by a Grant from the U.S. Department of Health, Education, and Welfare, Office of Education

SOUND FILM RECORDINGS IN IMPROVING CLASSROOM COMMUNICATIONS

Nathan Maccoby
Jon Jecker
Henry Breitrose
Ernest Rose

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION AND BACKGROUND	
1.1	Problem	1
1.2	Related Research	2
1.3	Goals	2
1.4	Assessment	3
1.5	Improvement	5
1.6	Coding and Training	8
II.	THE ASSESSMENT OF ACCURACY	
2.1	Introduction	11
2.2	Goals	12
2.3	Hypothesis	13
2.4	Preliminary Filming	13
2.5	Constructing a Scale	14
2.6	The Experimental Film	15
2.7	Figure 1: Original Scale for Judging Student Comprehension	16
2.8	Figure 2: Reproduction of Revised Student Comprehension Rating Scale and Supplementary Scales Used in the Main Experiment	17
2.9	Experimental Method	18
2.10	Table 1: Order of Presentation of Clips in Film	19
2.11	Experimental Design	20
2.12	Procedure	20
2.13	Table 2: Experimental Design	21
2.14	Results and Discussion	24
2.15	Table 3: Mean Number of Clips Rated Correctly	25
2.16	Table 4: Analysis of Variance of Number of Clips Rated Correctly	26
2.17	Table 4A: Mean Number of Clips Rated Correctly -- t-Comparisons Among Cell Means	27
2.18	Table 4B: Mean Number of Clips Rated Correctly -- t-Comparisons Between Cell Means and Means Expected by Chance	28
2.19	Table 5: Mean Number of Clips Rated Zero (can't tell) on Question A	30

CONTENTS (Continued)

Chapter		Page
II.	Continued	
2.20	Table 6: Analysis of Variance of Mean Number of Clips Rated Zero (can't tell) on Question A . . .	31
2.21	Table 7: Mean Confidence Ratings (Question B) . . .	33
2.22	Table 8: Analysis of Variance of Mean Confidence Ratings (Question B)	34
2.23	Summary	35
2.24	Table 9: Mean Effort Perceived Over All Clips Not Rated Zero (can't tell) by S	36
2.25	Table 10: Analysis of Variance of Mean Effort Perceived over all clips not rated zero (can't tell) by S. . .	37
III.	THE IMPROVEMENT OF ACCURACY	
3.1	Introduction	39
3.2	Method	39
3.3	Subjects	40
3.4	Collecting the Film Materials	40
3.5	The Pre-Training Test	42
3.6	Isolating Useful Non-Verbal Cues	44
3.7	The Training Procedure	45
3.8	Table 11: Items Analyzed From Code Form "Mark VII"	46
3.9	The Post-Training Test	47
3.10	The Effects of Training	49
3.11	Table 12: Comparison of Experimental and Control Pre-Training Mean Per Cent Accurate Scores Broken Down by Own Versus Others' Students	52
3.12	Table 13: Post Training Per Cent Accurate Scores Broken Down by Previous Exposure	54
3.13	Validity of Non-Verbal Cues as Predictors of Student Comprehension	55
3.14	Table 14: Partial Correlations of 15 Predictors of Student Performance with Actual Student Performance	57
IV.	PROBLEMS OF SOUND FILM RECORDING IN HIGH SCHOOL CLASSROOMS	
4.1	Recording the Visual Image	61
4.2	Recording the Sound	65
4.3	Sound Film Recording in Laboratory Situations . . .	70
4.4	Figure 3: Laboratory Filming Procedure	73
V.	SUMMARY	74
VI.	PERSONNEL	77

I. INTRODUCTION AND BACKGROUND

1.1 PROBLEM

To what extent can sound film recordings of pupils in teaching-learning situations be used as a basis for training teachers to improve their accuracy in interpreting non-verbal communications from students? Can visual cues from an audience be read accurately by a live communicator so that he can tell to what extent, if any, he is communicating successfully?

Communicators must keep on making judgments during the actual process of communication on how successfully they are communicating. Normally, in communications through mass media, this feedback is perforce delayed until audience research results become available. However, in face-to-face audio-visual communication settings such as in the classroom, the communicator (the teacher) does have information feeding back to him from the audience during all of the time that he is communicating. Presumably, he can keep on making judgments about whether or not he is being understood, whether he is being agreed with or disagreed with, and whether or not he is evoking strong emotion. In a two-person setting, immediate and recurring verbal feedback is frequently possible. When a small group is the audience, at least occasional feedback may take place although even here feedback is likely to be limited to particular audience members. When a teacher is addressing a sizable class of students, such verbal feedback is often severely limited. Aside from delayed feedback from test papers or project reports, the teacher is dependent largely upon facial expressions and other body movements of the

students for information on the effects and status of his communications.

1.2 RELATED RESEARCH

Previous research in the coding of facial expression by the students of kinesics (notably the work of Birdwhistell, Sebeck, Kramer and others who built on Sapir's original formulations) proved interesting but inappropriate for the scaling of cognitive rather than affective cues. The reliability and validity of such coding to our knowledge has never been previously established. Other work in methods of assessing non-verbal information in psychiatric interviews (such as that of Ruesch & Bateson) also proved to be not directly applicable to our problem.

1.3 GOALS

Because we suspected that visual feedback of the type described above is much depended upon, and because we suspected further that it is highly inaccurately interpreted by communicators, and because we thought that the accuracy of teachers in interpreting such non-verbal cues in the classroom could be improved, we decided to study it. We have been proceeding in our investigation of non-verbal feedback to teachers in the following manner. Using variable focal length camera lenses focused on the faces of individual students while the teacher is explaining something, we have been taking sound motion picuture recordings in schools with regular classes in session. One of the problems that we did not succeed in solving in the classroom itself originally was getting teachers to vary enormously in the success or failure with which they communicated something. We needed to get a great

deal of excursion in the degree of understanding among students in order to study differences in their non-verbal behavior under conditions of high cognitive clarity and under conditions of considerable confusion. For this reason, in addition to our classroom recordings, we brought individual students into our studio and, with the principal investigator serving as the teacher, we were able deliberately to achieve considerable variation in the clarity of the explanation so that we could see what the non-verbal behavior of students was under varying conditions of explanatory clarity.

1.4 ASSESSMENT

Our next problem was to see to what extent teachers could successfully read non-verbal feedback from students. We suspected that teachers rely heavily on such non-verbal feedback from students as a means of ascertaining the degree of comprehension students attained. We suspected further that teachers are not very accurate in the use of such cues. In a controlled experiment, we succeeded in demonstrating, at least in secondary schools, that teachers are indeed not very accurate in the interpretation of visual feedback from students. Furthermore, some of the best available current training of secondary school teachers apparently does not improve the teachers' ability to interpret such information. Nor does teaching experience do so.

Sound film records we have accumulated thus far suggest that there are cues that can be used more accurately to interpret non-verbal feedback than are currently employed. One of the first things we discovered was that there tends to be less similarity from person to person in the type of visual behavior used to communicate degree of

comprehension than is the case for verbal language. That is, when people within the same culture communicate verbally they can communicate successfully, at least within limits, because to a large extent they use what is a common language. However, when people communicate through facial expressions or body movements, the language is not necessarily so common within the culture. Perhaps one reason for this is that in the case of verbal communication, immediate verbal feedback is normally possible. Certainly in a one-to-one face-to-face situation this is so. Such immediate feedback serves to correct communication failures and leads, therefore, to a common language. This is not nearly as possible in the case of visual or non-verbal communication. Nevertheless, our explorations suggested there are at least intra-personal consistencies and when we get to know individual people well, we are more successful in reading their non-verbal behavior. Aside from this, however, there does seem to be an indication that there are some common inter-personal visual cues. For example, examination of film records suggests that the eyes must be focused on the communicator or on the visual device he is using before communication becomes optimally efficient. This, then, is a necessary but not sufficient condition for successful communication. That is, when the eyes are not focused on the communicator or on something he is demonstrating, communication is less than perfect. Furthermore, it seems to be the case that communicatees must remain in a relatively fixed position during reception for communication to be successful. When communicatees are engaged in a high degree of movement, either body movement of facial movement, communication to them is not being very successful. These observations

are consistent with those of Faison, Rose and Podell,[1] whose audiences, when highly interested in films being presented, remained in relatively fixed positions with eyes focused on the screen. Their infra-red films revealed that audiences viewing very dull films tended to be highly fidgety. We suspected that there are many other such inter-personal cues that are common for communicatees.

1.5 IMPROVEMENT

Our next step was to see whether or not we could successfully train communicators, such as teachers, to be more accurate than they currently are in interpreting non-verbal feedback from students with respect to the degree of comprehension of the material presented.

In addition, we have tested the hypotheses: "When teacher judgments of student comprehension are based on visual (non-verbal) cues from students, misperceptions of student comprehension are likely to occur more frequently than when verbal feedback from students is used principally." This hypothesis, as well as one testing the effectiveness of current teacher training and experience in improving the skill of teachers in reading non-verbal feedback cues of the degree of student comprehension, were tested and are reported in Section 3. Briefly the results confirm the prediction strongly. The inclusion of verbal information makes a very considerable difference in how well teachers and teacher-trainees can accurately characterize how well students are understanding what is being taught them. In fact, when only visual information is provided, subjects do no better than chance. This

---

1. Faison, E. W. J., Rose, N., & Podell, J. E. A technique for measuring observable audience reactions to training films. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL - LN - 55 - 45, 1955 (b). U -- mimeo.

experiment, however, was conducted with films taken in a laboratory or studio setting and needs replication with film recordings taken in the schoolroom. With the improved microphone setup for verbal recording of student response that we now have, we can conduct such an experiment in the classroom. Furthermore, in this experiment, no objective criterion of understanding on the part of the student was available. Judgments of the researchers were used. We indeed did have the additional information of what went on before and after the action shown in any particular film clip took place. Nevertheless, we could not be sure in any particular instance whether or not the student understood what was being taught. We now have such an objective criterion.

We then undertook the analysis of individual students' faces while they were being taught. We know independently both from the sound and from test scores of students taken immediately after instruction approximately what degree of comprehension is taking place. Preliminary analysis suggests that at least three variables are important in the visual feedback process. Teaching may be viewed as involving teaching/information-seeking situations in which the teacher is attempting to aid the student to learn. While the teacher is communicating information, the student ideally is attending and listening to the communicator. He varies first in the fixity of his orientation towards the teacher. That is, his eyes may be focused on the teacher or on what the teacher is demonstrating, and his body orientation may be relatively constant. This postural attitude has a given duration and then changes. Preliminary analysis suggested that high fixity of such a postural attitude is a necessary condition for attention and for

understanding. When attention wanders for any reason, this postural pattern shifts. The eyes look away and move to other objects of fixation. The facial muscles contract and grosser facial movements occur. The rest of the body, too, engages in more movement.

A second visual set of variables involves the movement of the muscles governing the forehead. When difficult material is being presented, concentration appears actually to be related to a furrowing of the brow; lack of attention appears correlated with a smoother forehead.

The third class of such variables consists of the presence or absence of specific facial and head gestures. These include head nodding vertically in assent, the lighting up of the eyes (in reality this is contraction of facial muscles around the eyes since the eyes themselves cannot change expression), looks of puzzlement (again facial muscle pattern changes), and so on.

These considerations led us to attempt another experiment. This time we decided to see if using all the means at our disposal, we could improve the ability of teachers correctly to interpret visual feedback from students being taught. Our first problem was to achieve control over what was being taught in the classroom so that difficulty and clarity of instruction could be manipulated in standard fashion. To achieve this end, we adapted a programmed instruction in kinship systems, originally prepared by Professor A. K. Romney of Stanford's Anthropology Department and Professor D. E. Allen of the Stanford Secondary Education Project. Because we needed to take motion picture films of students being taught, we could not have the program in a book or

machine at each desk. Instead, we had teachers instruct and put multiple choice practice questions on answer sheets at their desks. The films taken thus consisted of a series of clips, where a clip consists of a shot of the student while being taught a specific item of information and includes his answer to a question based on that point. Typically, such clips are about thirty-seconds long but vary somewhat with the nature of the item being taught. Because the researchers have a record of each student's answers, we know whether or not the student in the clip understood what was being explained to him. To be sure, on occasion he may get the correct answer by chance, or he may not get it until his face is out of view because his head is down toward the answer sheet. But, by and large, we regard this procedure as a satisfactory answer to the above mentioned problem of establishing an objective criterion of student's understanding of what he is being taught.

1.6 CODING AND TRAINING

Our chief hypothesis in this experiment was that when teachers are taught to pay attention to relevant visual cues and to characterize these cues in objective terms, it will improve their ability to interpret these cues as indicating understanding in a pupil or the lack of it. It then became necessary to develop a reliable set of codes in order to categorize these visual imputs. These codes had to be objective ones; that is, consist of descriptions of particular facial and bodily movements if they were to be clear enough to permit the attainment of satisfactory coding reliabilities. They had to be sufficiently valid to improve prediction of understanding or its absence beyond a chance

base line and beyond a base line of initial performance without training if such initial performance by teachers were indeed better than chance.

Our best judgments of what such code categories should be were derived from our earlier analyses described previously. Such body and facial movements as orienting oneself towards the source of information, frequency and speed of movement to and away from such sources, brow furrowing and raising, chin rubbing, and the like, became the nuclei of such codes. We experienced tremendous difficulty in achieving adequate degrees of agreement among raters in code revision after code revision, until we applied some standard principles that have been developed for the categorizing of verbal content. Some of these principles are that code categories should be mutually exclusive, should be parallel, and should be exhaustive of the possible contents. Finally, we built a version of the code (Mark I) that met these criteria. Although our initial assays at agreement among independent coders varied and were not uniformly satisfactory, we had the means at our disposal for improvement. By comparing disagreements and re-running any given film clip, we could either redefine the category so as to sharpen it or discover problems in the observation process that could be trained for. After a long series of revisions based on such a procedure, we developed a version of the code that provided adequately reliable items (Mark VII).

The validity of the code is best measured by the extent to which it can be used to predict whether or not student comprehension occurs. To this end, we planned to carry out multiple regression analyses to discover what weights optimally to assign each code category in order to maximize successful prediction of comprehension. With the development

of a reliable code and of an objective measure of comprehension the experiment was designed and performed under controlled conditions. Results support the hypothesis that teacher accuracy in assessing cognitive visual feedback can be significantly improved by training.

II. THE ASSESSMENT OF ACCURACY[2]

2.1 INTRODUCTION

Communicators must continually make judgments during the actual process of communcation on how successfully they are communicating. Normally in communication through mass media, this feedback is perforce delayed until audience research results become available. However, in face-to-face communication settings, the communicator does have information feeding back to him from the audience during all of the time he is communicating. Presumably, he can keep making judgments about whether or not he is being understood, he is being agreed with or disagreed with, or whether or not he is evoking strong emotion. In a two-person setting, immediate and recurring verbal feedback is often possible. Even when a small group is the audience at least occasional feedback may take place, although even here feedback is likely to be limited largely to particular audience members. When a teacher is addressing a sizeable class of students, such verbal feedback is likely to be severely limited. Aside from delayed feedback from test papers or project reports, the teacher is largely dependent upon the facial expressions and other body movements of the students for information on the effects of his communication.

It was our suspicion that such visual feedback is much depended upon, and, that it is highly inaccurately interpreted by communicators.

Although there has been considerable previous research reported on interpretation of non-verbal cues from people, this research has primarily been concerned with the interpretation of affective information.

---

2. Jon D. Jacker, Nathan Maccoby, Henry S. Breitrose and Ernest D. Rose, Institute for Communication Research, Stanford University.

The present research is concerned entirely with cognitive non-verbal cues.

2.2 GOALS

Our intention, in the following experiment, was to conduct some preliminary and exploratory research investigating student-teacher communication. One of our stated project objectives[3] is "to discover the patterns of visual feedback from students to teachers in the classroom situation which are currently used or could be used as indices of the success or failure of student-teacher communication". Specifically, our concern in this experiment was to investigate the extent to which people, either teachers or those who are training to be teachers, could correctly judge how much a student is comprehending of the material being presented. Our concern stems chiefly from an interest in the use of non-verbal cues, specifically visual ones, as a basis for making such cognitive judgments.

In the attempt to achieve the objective stated above, the first step was to assess the relative value of verbal and non-verbal cues. If we could thus assess the extent to which verbal cues contribute to success in judging student comprehension, we hoped to establish, as a second phase of this research, some types of non-verbal cues which are given standard interpretations by people using them to judge student comprehension.

To our knowledge, no previous attempt has been made to establish the relative value of verbal and non-verbal cues in making judgments of student comprehension. This is due, in part, to the difficulty of

---

3. The research reported in this paper is part of NDEA Title VII project entitled "Sound film recordings in improving classroom communications."

discovering a technique for keeping a record of total feedback from student to teacher. Our feeling was that a sound film recording of actual student-teacher interaction resolved this difficulty rather admirably. The sound film recording technique would provide an extensive and direct record of both verbal and non-verbal behavior on the part of the student.

2.3 HYPOTHESIS

In the absence of any previous research on the frequency, nature, use and degree of teacher awareness of non-verbal cues, our hypothesis was based largely on intuition. Our specific hypothesis was as follows: When teacher judgments of student comprehension are based on visual (non-verbal) cues from students, misperceptions of student comprehension are likely to occur more frequently than when verbal feedback from students is used principally.

Our second chief interest in this study was to investigate differences which might exist between non-teachers, new teachers and experienced teachers in their ability to make accurate judgements. Any differences exhibited should be in line with experience and training. Specifically, we hypothesized that: any differences among these "experience groups" should be consistent with a positive relationship between experience and success in judging student comprehension.

2.4 PRELIMINARY FILMING

The first procedural step was to collect the sound film recordings of student-teacher interaction needed as stimulus material in a controlled investigation. This was done by making sound film recordings of a number of students being taught either elementary algebra or

elementary statistics by one of the investigators in our studio. Twenty (20) junior high school (8th or 9th grade) students were filmed while being taught. Figure 3 (page 73) gives the reader an idea of the actual arrangements of this filming set-up. It is worth mentioning here that the students experienced no difficulty in ignoring the camera and paying attention to the instructor. Students were obviously aware of being filmed, and of course suitably impressed, but were apparently not seriously distracted from the instruction. The situation appeared to resemble the performance of players in televised athletic contests. When play is not in action, players occasionally attend to the camera, but when play starts, this is rarely the case. Players become too busy.

The instruction in elementary algebra was given to individual students. Hopefully for purposes of economy, the later instruction in elementary statistics was given to pairs of students. Two cameras were used to obtain individual films of the students. This proved to be false economy, because in later editing the films for use as stimulus material in the experimental film, there was such a scarcity of verbal cues on the part of the students instructed in pairs that these films proved to be of no use for the purposes of this experiment. They are of value to us, of course, in analyzing non-verbal cues. Such a scarcity of verbal cues prevented adequate comparison of the relative value of verbal and non-verbal cues. For this reason, only the films of instruction in elementary algebra to individual students were used as stimulus material in the experiment.

2.5 CONSTRUCTING A SCALE

Since we wanted to investigate the extent of accuracy in judging

the amount of student comprehension in segments of student-teacher interaction, our first experimental step was to determine, via pretest, a usable scale for making such judgments. To implement this pretest, the experimenters designed what appeared to be a reasonable scale (Figure 1, page 16) for making such judgments, and chose five segments, or "clips", from the film of each of four students. The criteria for choosing each clip was that it be a recognizable question-answer sequence and that it fit one of the categories of the previously determined scale of student comprehension. This resulted in a series of twenty clips, which were then spliced together, separated by blank leader, for purposes of presentation.

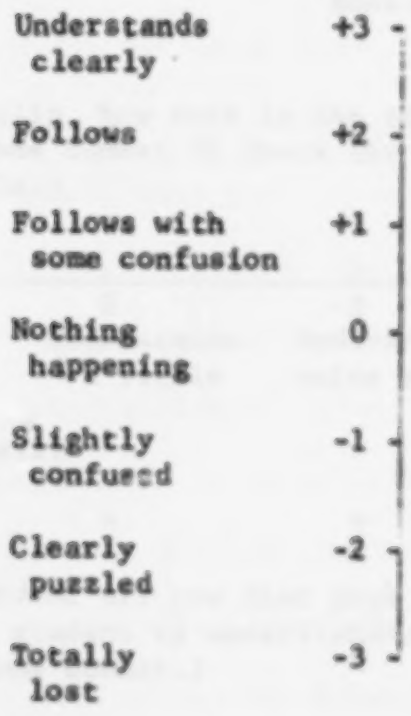
Twelve (12) pretest subjects were shown this twenty clip film, and were asked to use our rating scale to judge the amount of student comprehension in each clip, immediately after it was shown. Subjects were encouraged to discuss the basis of their ratings. After using our scale to rate each of the twenty clips, they were interviewed by the experimenter as to what modifications should be made in the scale to improve its utility in this setting. These interviews were tape recorded and later content analyzed to determine what revision of the original scale to actually carry out. The revised scale, which was the only noteworthy result of the pretest, is presented in Figure 2 (page 17). (Question A)

2.6 THE EXPERIMENTAL FILM

Using the revised scale for rating student comprehension which resulted from the pretest, the investigators first determined what comprehension-category clip to take from each of the ten usable instruction

2.7 FIGURE 1

Original scale for judging student comprehension (used in pretest).



2.8 FIGURE 2

Reproduction of revised student comprehension rating scale and supplementary scales used in the main experiment.

Clip #____ Booklet # 0000

A. In this clip, how much is the student understanding? (Circle one number OR check the statement below the scale.)

1	2	3	4
Understands nothing	Understands a little	Understands quite a bit	Understands completely

____Can't tell.

* * * *

B. How confident are you that your judgment of how much the student is understanding is correct? (Circle one number.)

1	2	3	4	5
Not at all confident				Totally confident

* * * *

C. In this clip, how much GENUINE effort is the student making to understand? (Check only one.)

____A great deal of effort.

____Some effort.

____Very little effort.

____No effort at all.

____Can't tell.

films. Four of these films were of female students, six of male students. Two clips were taken from each film. The order of presentation of individual students, and the comprehension-category (amount of understanding) selected from the film of each student were each randomly determined. The order of presenting sexes was alternating, except for some adjacent male clips due to the six male - four female imbalance. Table 1 (page 19) shows the complete order of presentation used in putting together the experimental film.

After pre-determining what type of clip to pull from each of the available films, clips were chosen on the basis of being the best example of each comprehension-category in their respective films. The editors involved used all possible contextual information to rate and select the clips. Only clips on which one hundred percent editor agreement was achieved were used in the final experimental film. Since the rating categories were broad, the editors felt quite confident of their own accuracy in selecting the clips.

The experimental film was divided into two sets of ten clips, with each instruction film supplying one clip in each set.

2.9 EXPERIMENTAL METHOD

Subjects

Subjects were scheduled with the cooperation of the Secondary Education Project in the School of Education at Stanford. Subjects were recruited from three experience levels. Sixty-seven (67) Stanford students who had just enrolled in the teacher training program at Stanford made up the least experience, or NEW INTERN group. Fifty-nine (59) Stanford students who were nearing completion of the same teacher training

2.10 TABLE 1

Order of presentation of clips in film.

Clip #	Sex of student	Our rating
1	M	2
2	M	3
3	F	1
4	M	0
5	F	4
6	M	1
7	F	3
8	M	4
9	F	0
10	M	2
11	M	2
12	M	3
13	M	1
14	F	0
15	M	4
16	F	1
17	M	3
18	F	4
19	M	0
20	F	2

program but who had not yet taken their first paying job as a teacher made up the moderate experience, or OLD INTERN group. Forty-six (46) teachers, enrolled in summer courses at Stanford and having at least two years teaching experience made up the highest experience, or EXPERIENCED TEACHER group. The total N was thus one hundred seventy-two (172).

2.11 EXPERIMENTAL DESIGN

Each experience level group was subdivided into three experimental conditions. Approximately one third of each experience level group was shown the whole experimental film, i.e. both the picture and the sound tracks. This PICTURE PLUS SOUND CONDITION will hereafter be referred to by PIX+SD. Approximately one third of each experience level group heard only the sound track (SOUND ONLY CONDITION designated by S-O), and the remaining subjects in each experience level group saw only the picture track of the experimental film (PICTURE ONLY CONDITION designated by P-O). Table 2 (page 21) presents the experimental design with N for the various cells shown.

2.12 PROCEDURE

The procedure called for running the experiment in six groups. Proportionate numbers of subjects from each experience level group were pre-assigned to each experimental hour, and all subjects in one experimental hour were in the same condition, PIX+SD, S-O or P-O. As is obvious from Table 2, the relationship between sign-up and show-up was less than perfect.

When subjects arrived at the experiment, they were handed a printed booklet. The first page of this booklet was devoted to a photograph of

2.13

TABLE 2

Experimental Design

	Picture plus Sound	Sound Only	Picture Only	Marginal
New Interns	N=20	27	20	67
Old Interns	N=20	18	21	59
Experienced Teachers	N=17	15	14	46
Marginal	N=57	60	55	Total N 172

the setting in which the instruction films were taken. This was the same photograph as Figure 3 of this report. Pages two through twenty-one were identical, and contained the questions with their respective rating scales to be answered in response to each clip of the experimental film. These questions and scales are presented in Figure 2.

Instructions to the groups were brief and to the point. A short explanation of the investigators' interests was given. A familiarization with the setting in which the films were taken was achieved with the help of the photograph on the first page of the booklet. The various rating scales in the booklet were discussed in detail and instructions were given concerning the manner of answering each, complete with blackboard examples, in the best tradition of classroom instruction. Instruction on answering Question A included the strong suggestion that the "Can't Tell" category be used very little. Subjects were admonished to rate the degree of student comprehension whenever possible.

Instruction on answering Question B emphasized that a confidence rating should be given in all cases, even when Question A had been answered "Can't Tell". In these cases, they were instructed to rate how confident they were that no cues for rating student comprehension were present.

Instruction on answering Question C consisted of reading that question and contrasting "genuine effort" with attempting to give the impression of paying attention while actually thinking on unrelated topics. Our reason for asking Question C stemmed from the pretest results. Content analysis of the pretest data revealed that judgments of

student comprehension were, in some cases, confused by the raters' perceptions concerning how hard the student in the film was trying to understand. This confusion generally resulted in judging more student comprehension where high student effort was perceived.

Following these instructions, and any questions which arose, two example film-clips were shown; and following each, the experimenter marked duplicates of all three rating scales at the blackboard. The example film-clips were shown complete, i.e. picture track and sound track, in all groups.

When the example clips had been dispensed with, the experimenter then instructed the group concerning which experimental treatment they were to receive, and the first ten clips of the experimental film were shown. A thirty second interval was interposed after each clip during which subjects were to make their ratings. The room lights were turned off during the presentation of the clips, even in the S-O condition. Lighting was restored during each thirty second rating period.

At the end of the first ten clips, room lights were left on and the experimenter suggested that subjects take a break and complete the questionnaire placed loose in the middle of the booklet. Information obtained on this questionnaire served to identify each subject and place him in one of the experience level groups. Also, birth order data was collected. The rest of the questions served simply as filler material to keep subjects busy and prevent discussion of the experimental task. The experimenter also explicitly requested that subjects not discuss their ratings during this period. Approximately five minutes were allowed each group for this break.

The second set of ten clips was shown with no additional instructions. After the last clip, the lights were left on, booklets were collected, and subjects were dismissed.

2.14 RESULTS AND DISCUSSION

The main result of the experiment concerns the data supplied by the answers to Question A (Figure 2). Using the ratings preassigned by the investigators as an accuracy criterion, each subject was given an accuracy score based on the number of clips judged correctly. Table 3 (page 25) presents the mean number of clips judged correctly in the nine cells of the experimental design. Table 4 (page 26) presents the analysis of variance of these data. As can be seen, a highly significant experimental treatment, or Information Level effect is shown ($F = 105.6$, $p < .001$), with the PIX+SD and S-O cells supplying means of significantly greater accuracy than the P-O cells (Table 4A, page 27). PIX+SD and S-O means are not significantly different from each other, and both indicate significantly greater accuracy than could be expected on the basis of chance performance (Table 4B, page 28). The P-O mean, on the other hand, does not show significantly greater accuracy than would be expected on the basis of chance.

The analysis of variance and related t-tests provide no support for an hypothesis that experience making judgments identical in content to those in the experiment improves the accuracy of those judgments. There are no differences in accuracy between experience groups which even approach statistical significance. There is not even a significant interaction to suggest that experience exerts an influence in some of the experimental conditions.

2.15 TABLE 3

Mean number of clips rated correctly (Question A)

	Picture plus Sound	Sound Only	Picture Only	Marginal
New Interns	9.5	11.4	5.3	9.0
Old Interns	10.8	11.4	5.2	9.0
Experienced Teachers	10.3	10.2	4.8	8.6
Marginal	10.2	11.1	5.1	

2.16 TABLE 4

Analysis of variance of number of clips rated correctly (Question A)

Source	F	P
Information	105.6	<.001
Experience	1.2	ns
Interaction	1.2	ns

2.17 TABLE 4A

Mean number of clips rated correctly - t-comparisons among cell means. (Question A)

Comparison		t	P
New Interns	P+S > P-0	6.26	<.001
	S-0 > P-0	8.17	<.001
	P+S > S-0	2.60	<.02
Old Interns	P+S > P-0	8.55	<.001
	S-0 > P-0	7.96	<.001
	P+S ≠ S-0	<1	ns
Exp. Teachers	P+S > P-0	6.23	<.001
	S-0 > P-0	9.30	<.001
	P+S ≠ S-0	<1	ns

2.18 TABLE 4B

Mean number of clips rated correctly - t-comparisons between cell means and mean expected by chance (4.0).

Comparison		t	P
New Interns	P+S > 4.0	2.62	<.02
	S-O > 4.0	2.64	<.02
	P+O > 4.0	<1.00	ns
Old Interns	P+S > 4.0	3.09	<.01
	S-O > 4.0	2.74	<.02
	P+S > 4.0	<1.00	ns
Exp. Teachers	P+S > 4.0	2.1	=.06
	S-O > 4.0	4.13	<.01
	P+O > 4.0	<1.00	ns

The interpretation of these data is clear. First, our hypothesis concerning the relative value of verbal and non-verbal cues is confirmed at an unquestionable level of significance. Verbal cues produce far greater accuracy than non-verbal cues. The latter seem not to even contribute to accuracy when they occur simultaneously with verbal cues. Secondly, people required to make this type of judgment seem not to improve their accuracy through ordinary teacher training or through experience in classroom teaching.

An interesting result is found in looking at the number of clips rated zero, or "Can't Tell" on Question A. Table 5 (page 30) presents the mean number of clips rated zero in each cell of the design. Table 6 (page 31) presents the analysis of variance on these data. This analysis makes it clear that the condition in which subjects are least accurate (P-0) is the one in which the"Can't Tell" category is least often used. The investigators feel that the "Can't Tell" category is frequently used in the S-0 condition because there is a marked contrast between the clarity of verbal cues in most clips and the marked scarcity of verbal cues in the remaining clips. Consistent with this contrast hypothesis, the scarcity of clear, useful cues in nearly all the clips in the P-0 condition leads S to the feeling that if he rates amount of comprehension in one clip, he may as well do the same with all, or nearly all clips. There is no apparent contrast in the amount of useful cues present in the clips when only the picture track is presented.

In reflecting on the strength of verbal cues as they effect accuracy of judgments, it became apparent to the investigators that a related factor - silence - is also highly significant. To check this

2.19 TABLE 5

Mean number of clips rated zero (Can't tell) on Question A.

	Picture plus Sound	Sound Only	Picture Only	Marginal
New Interns	3.0	4.6	1.0	3.0
Old Interns	2.8	5.4	1.3	3.1
Experienced Teachers	2.8	4.0	0.8	2.6
Marginal	2.9	4.7	1.1	

2.20

TABLE 6

Analysis of variance of mean number of clips rated zero (Can't tell) on Question A.

Source	F	P
Information	45.1	<.001
Experience	1.4	ns
Interaction	0.6	ns

silence effect, the following analysis was performed. The latency between the first posing of a question by the instructor and the answering response by the student was determined for fourteen of the twenty clips. The remaining six clips were not amenable to such an analysis since they did not contain a student response sufficiently specific to demarcate the question-answer sequence.

These fourteen scorable clips were then arranged in order of their latencies and dichotomized on that dimension. A Fisher Exact test of the criterion ratings of the clips in these two latency categories results in $p = .0023$. To interpret, the investigators show a very strong tendency to judge understanding in an inverse relationship to latency between question and answer.

It is very difficult to eliminate latency as a basic cue for interpretation of degree of comprehension because latency really includes not just the silent period preceeding a student's reply but also pauses and hesitations while the reply is being made.

An analysis of the confidence rating data is also of interest. Each subject was given an average confidence score over all twenty clips. Table 7 (page 33) presents the mean confidence in each of the nine cells of the design. Table 8 (page 34) presents the analysis of variance of these data. Again a significant experimental treatment effect is shown ($F = 6.71$, $p = <.01$), with no other significant effects. It is reassuring to see that where accuracy is not significantly greater than chance, i.e. the P-O condition, confidence in ratings is low.

Another analysis performed was on the data supplied by Question C, the subjects perception of the amount of effort being exerted by the

2.21 TABLE 7

Mean confidence ratings (Question B).

	Picture plus Sound	Sound Only	Picture Only	Marginal
New Interns	3.59	3.73	2.94	3.38
Old Interns	3.73	3.06	3.30	3.20
Experienced Teachers	3.86	3.40	3.11	3.45
Marginal	3.54	3.38	3.09	

2.22 TABLE 8

Analysis of variance of Mean confidence ratings (Question B).

Source	F	P
Information	6.71	<.01
Experience	2.30	ns
Interaction	1.95	ns

student in the clip to understand the material being presented. Table 9 (page 36) presents means, and Table 10 (page 37) presents the analysis of variance on these data. The only purpose in including the question was to remove any bias, by all subjects which perceived effort might have on judgments of amount of student comprehension. Because such a bias would tend to reduce accuracy, it seemed desirable to minimize it. No hypothesis concerning differences among conditions were advanced, since such a bias was presumed to operate in all groups, It might be interesting to speculate about the significant Information Level effect demonstrated in these data, but that is beyond the scope of this paper.

The only other data analysed were those on birth order effects. and no significant effects could be attributed to birth order as a factor, based on the analyses performed.

2.23 SUMMARY

An experiment was conducted to investigate the relative value of verbal and non-verbal cues as they contribute to the accuracy with which teachers make judgments concerning student comprehension of the material being presented. Non-teachers, inexperienced teachers, and experienced teachers were shown sound film clips of student-teacher interactions, and were asked to rate the amount of student comprehension evident in the clips. Some subjects were shown only the picture track, some heard only the sound track, and some saw the complete sound film recording. The results strongly indicate that verbal cues contribute significantly to accuracy of judgment, while the presence of non-verbal cues, specifically visual ones, do not increase accuracy of judgments above chance expectancy. Also, no evidence was found to

2.24 TABLE 9

Mean effort perceived over all clips not rated zero (Can't tell) by S.

	Picture plus Sound	Sound Only	Picture Only	Marginal
New Interns	2.94	2.99	3.00	2.98
Old Interns	2.86	3.00	3.01	2.96
Experienced Teachers	2.98	2.91	3.25	3.04
Marginal	2.92	2.97	3.07	

2.25 TABLE 10

Analysis of variance of mean effort perceived over all clips not rated zero (Can't tell) by S.

Source	F	P
Information	4.1	<.05
Experience	1.1	ns
Interaction	1.8	ns

support the hypothesis that normal teacher training or teaching experience making content-identical judgments improve the accuracy of those judgments. It was shown that low accuracy was accompanied by low confidence in the judgments made.

III. THE IMPROVEMENT OF ACCURACY

3.1 INTRODUCTION

A remark frequently heard from students as they leave a classroom goes something like this: "I was lost after the first 5 minutes." The implication is clear that the student failed to understand the bulk of the lesson because some early point integral to the lesson was not understood. Similarly, teachers frequently belabor a point long after students have grasped it. In general, it is a difficult task to gear classroom teaching to achieve optimal student comprehension. This problem could be largely avoided if the teacher were able to make accurate judgments of student comprehension at various critical points during the lesson.

The present investigators, have shown that such judgments can be made with significant accuracy when verbal feedback from the students is available, but found no evidence that such non-verbal feedback is being currently used successfully as cues of student comprehension.

It was our belief that non-verbal feedback should provide some useful cues for making such judgments. The following is a report of an attempt to discover useful non-verbal cues and train teachers to use them.

3.2 METHOD

Sound-film recordings were made of individual students while they were being taught by their regular instructors. Due to the nature of the lesson being taught, which will be discussed in detail below, the investigators were able to isolate short film segments, "clips," and score each for student comprehension. A test film, consisting of a large

number of these clips, was then shown to each teacher. The teacher's task was to make a judgment of student comprehension in each clip.

The teachers were then split into experimental and control groups. Experimental teachers viewed many additional clips, and they were trained to relate non-verbal cues to student comprehension. Control subjects received placebo training during which they saw and discussed commercially available educational films dealing with the general area of interpersonal communication.

A second test film, similar to the first one, was then administered to all subjects.

3.3 SUBJECTS

Twenty graduate students doing their supervised teaching under the direction of the teacher training program in the School of Education at Stanford University served as subjects.[4]

3.4 COLLECTING THE FILM MATERIALS

Since it was necessary to collect films in 20 different classrooms during instruction by 20 different teachers, it seemed desirable to reduce the variability of content of some of the lessons. Further, since the films were to be used as test materials for judging student comprehension, we needed some method of independently assessing the degree of actual student comprehension. To satisfy these two needs, all teachers were provided with a standard "lesson" consisting of 35 items of instruction in the field of anthropology, on the concept of familial structure. Generally, this material was new to the students,

---

4. One of these had just completed this program and had taken his first teaching job at the time of the experiment.

and at a suitable level of difficulty. The lesson was taught verbally by the teacher in a manner not unlike normal classroom teaching technique, and immediately after each item was presented, the teacher posed a question to the students concerning that item. The students were provided with answer sheets and recorded their own answers by checking the appropriate answer.

Each question was presented on large cards mounted on an easel at the front of the class. Students were requested to reserve any questions of the teacher until the completion of the lesson, thus minimizing verbal feedback from students to teacher. Each question card had a set of multiple-choice type answers, the last of which was "I don't know." Students were instructed to answer each question as best they could, and to use the last answer whenever they <u>really</u> had no idea which was the correct answer. As the teacher finished each item of instruction, she would reveal the question card concerning that item and read the question. She would pause 20-30 seconds for the students to answer the question, and then step in front of the question card to present the next instruction item. This procedure for presenting questions helped to maximize the amount of time during which we could photograph the students' faces.

Two 16 mm. Auricon cameras with 33 minute film magazines were used to obtain continuous photographs of individual students as they responded to the lesson. The cameras were equipped with variable focal length lenses, so that the image size of the student on the film could be kept reasonably constant despite variation in distance from camera to student. Cameramen were instructed to pick individual students in the

class and photograph them for a period of either 5 or 10 instruction items. (The purpose of this variation will become apparent during discussion of the construction of the posttest.) A separate synchronous sound recording was made of the teacher's instruction.

These films were then broken down into "clips." A clip consisted of the film footage on one student from the beginning of one instruction item up to but not including the beginning of the next item, usually about 45 seconds. Each clip thus included the period of time during which the student was expected to answer the question for that item. Through the use of the synchronous sound recording, each clip on a student could be identified as to the specific instruction item presented during that clip. By consulting the student's answer sheet, each clip was scored "right" or "wrong" depending on whether the student answered the question on that item correctly or incorrectly. Thus each clip was identified by teacher, student, item number, and right or wrong answer. Once identified, the clips could be assembled into test films in any order desirable.

3.5 THE PRE-TRAINING TEST

The first experimental objective was to assess the ability of the teachers to judge student comprehension accurately before any training. To this end, four pre-training test films were constructed in the following manner: From the films of each teacher's class, 25 clips, distributed evenly over 5 students, were drawn.[5] These clips were drawn from students filmed for 10 consecutive items so that a post-training

---

5. Due to incomplete films, less than 25 clips were drawn for six teachers. For these six, we drew 24, 24, 24, 20, 15, and 13 clips.

test could include new clips on these same students. Since the instruction and accompanying questions grew increasingly difficult during the lesson period, each set of five clips on any one student was randomized to eliminate item difficulty as a clue in judging student comprehension.

Groups of four teachers were assigned to five pre-training test groups. Approximately 20 randomized sets of five clips on single students were presented in each pre-training test film by following a fixed order of teachers to which the student in each set belonged. Thus each teacher saw one set on one of her students, and then three sets, one of which belonged to each of the other three teachers in the test group. This fixed order of five-clip sets was repeated throughout each test film.

The test films were shown silent for two reasons: First, we were interested in judgments based on non-verbal cues only. Second, while the sound track contained only the teachers' instruction, it could provide cues concerning the order and thus the difficulty of each item. Before beginning the test, all teachers were instructed that each clip they would see would show a student during the instruction of one item and the time allotted to answer the question on that item, and that the clips would not necessarily be shown in the order in which they were photographed. They were instructed to judge whether the student answered the question in each clip correctly or incorrectly, and were to judge all clips, those of their own students and those of others'. An answer sheet was provided for this purpose. Each teacher made from 84 to 100 such judgments during the pre-training test.

3.6 ISOLATING USEFUL NON-VERBAL CUES

The films were next viewed by the investigators in an effort to select non-verbal cues which might prove useful in judging student comprehension. The films allowed us to investigate many relationships between specific non-verbal behaviors and the student's actual performance on the question answered for each clip. Facial movements and expressions, as well as various other specific behaviors of a non-verbal nature were considered for their utility in assessing comprehension. After many revisions the investigators agreed on a set of 22 specific non-verbal cues as being related to comprehension. For each of these cues, a set of scoring categories was designed so that a score could be given on every cue for any clip. The cue items were then assembled into a single-sheet coding form which could be completed for any and every clip viewed. The completed cue-coding form resembled those used in standard interviewing techniques. Such body and facial movements as orientation toward the source of information, frequency and speed of movement to and away from such sources, brow furrowing and raising, chin rubbing, and the like, became the nuclei of such codes. We experienced considerable difficulty in achieving adequate degrees of agreement among raters in code revision after code revision, until we applied some standard principles that have been developed for the categorizing of verbal content. Some of these principles are that code categories should be mutually exclusive, should be parallel, and should be exhaustive of the possible contents. Finally, we built a version of the code (Mark I) that met these criteria. Although our initial assays of agreement among independent coders varied and were not uniformly satisfactory,

we had the means at our disposal for improvement. By comparing disagreements and rerunning any given film clip, we could either redefine the category so as to sharpen it or discover problems in the observation process that could be trained for. After a long series of revisions based on such a procedure, we developed a version of the code that provided adequately reliable items (Mark VII -- see Table 11, page 46). The first item is an identifying number. The last item is a judgment of student comprehension. Some of the items of this code could still stand some improvement in reliability, but on the whole, agreement ratios approximated those normally attained in the coding of open-ended verbal material. Coding reliability was measured in the usual way; in terms of the ratio of agreements among independent coders coding the same material to the total number of possible agreements. When the data, consisting of film clips, were coded independently by either four or five members of the research team, such ratios ranged from a low of 66.8 per cent agreement (where 40 per cent would be the chance base line) to 98.7 per cent agreement (where 50 per cent would be chance). The over-all mean ratio was 78.5 per cent. Of course, some of the categories are quite easy to score reliably in that the behavior being coded occurred only rarely, and coders could easily agree that it did not occur in a particular clip. Such items were retained, however, because when such behavior does occur, it may be highly indicative of the degree of comprehension.

3.7 THE TRAINING PROCEDURE

The 20 teachers were next assigned to one of two groups by the toss of a coin. One restriction was placed on this assignment: males

3.8

TABLE 11

Items Analyzed From Code Form "Mark VII"

1-Clip Number _____

2-Amount of time looked at source
1-Little or none
2-Somewhat
3-A great deal
4-Virtually all the time

3-Number of times looked away from source
1-Many times (more than 4)
2-Some (3 or 4)
3-Little (1 or 2)
4-None

4-Speed of eye movements away from and returning to the source
1-Slow
2-Fast
3-None away from source

5-Amount of blinking
1-More than normal
2-Normal
3-Less than normal

6-Grouping of blinking
1-Spaced
2-Clustered
3-Virtually no blinking

7-Duration of lowering eyebrows
1-A lot
2-A little
3-None

8-Strength of lowering eyebrows
1-Strong
2-Weak
3-No eyebrow lowering

9-Duration of raising eyebrows
1-A lot
2-A little
3-None

10-Strength of raising eyebrows
1-Strong
2-Weak
3-No eyebrow raising

11-Movement of hands on face
1-A lot
2-Some
3-A little
4-None

12-Frequency of general body movement
1-A lot
2-Some
3-A little or none

13-Extent of general body movement
1-A lot
2-Some
3-A little or none

14-Amount of mouth movement
1-A lot
2-Some
3-A little or none

15-Chewing gum or candy
1-Yes
2-No

16-My guess of students answer
1-Wrong
2-Right

17-Students actual answer
1-Wrong
2-Right

and females were distributed as evenly as possible between the groups. One of these groups was designated as the experimental training group. Their training procedure was as follows: Four separate sessions of 1-1/2 to 2 hours each were conducted. At each training session, teachers viewed from 15 to 30 clips, completing a coding form on non-verbal cues for each clip. Open discussion occurred in all training sessions as to the correct scoring, interpretation and meaning of non-verbal cues. After coding the non-verbal cues in each clip, teachers judged the student in the clip as answering correctly or incorrectly, and were then given feedback on the student's actual performance.

The other group of ten teachers was designated as the control group. They also met for four sessions of 1-1/2 to 2 hours. During each session, they were shown a film concerning interpersonal communication and discussed those aspects of each film which were relevant to the student-teacher communication process.

3.9 THE POST-TRAINING TEST

Following the completion of the four training sessions, the teachers were assigned to seven post-training test groups, six with three teachers and one with two teachers. Seven new test films were prepared, each containing 90 clips. These test films were constructed so as to contain clips of three types: 30 repeats of clips shown in the pretest and used in experimental training sessions, 30 clips on students shown in the pretest clips which did not themselves appear in the pre-training test or in training; and 30 new clips on students not shown in the pre-training test. Some variation in the exact number of clips of each type occurred due to incomplete films in some classes. Those students

who had been photographed for ten consecutive items supplied clips in each of the first two of the above types. Students photographed for five consecutive items supplied clips in the last category. These three types of items were deliberately included to check on two types of practice effects which might occur in the experimental training group. It might occur that teachers in the experimental condition simply learned, from feedback in the training sessions, the correct judgments for some of the clips used in training. A comparison of improvement in judging clips of types I with clips of types II and III provides a check on such a practice effect. It might also occur that the teachers in the experimental condition may have learned to judge the students shown during training. A comparison of improvement in judging clips of types I and II with clips of type III provides a check on this kind of practice effect.

The clips were presented in the post-training tests in pairs of clips of each student. The order of presentation of these pairs was again fixed so as to rotate through the teachers viewing each test film. The order of original occurrence of the items was not randomized within each pair of clips, but the order of the type of clip was randomized within each set of three pairs, i.e., within each rotation through teachers. Thus each teacher saw one set of two clips on one of her students, and then two additional sets, one of which belonged to each of the other teachers in the test group; and each set of three pairs of clips contained one pair of each type of clip.

Completion of the post-training test marked the end of the experiment.

3.10 THE EFFECTS OF TRAINING

The pre-training test scores were analyzed first. Each teacher was given a "per cent accurate" score by dividing the number of correct judgments of student comprehension by the total number of judgments made. As mentioned above, our previous research found no evidence that visual cues were being successfully used to judge student comprehension. The result obtained on the pre-training test in this study is consistent with the earlier finding. The mean per cent score obtained here for all twenty teachers was 56.7 per cent accurate. This is not significantly different from a chance accuracy score of 50 per cent ($t = 1.12$).

It will be recalled that following the pre-training test, teachers were assigned by the flip of a coin to experimental and control groups. Ideally, performance on the pre-training test should not differ between the two groups. Unfortunately, comparison of these scores reveals that teachers assigned to the control group had a significantly higher mean per cent score (59.8 per cent) than did the experimental teachers (53.7 per cent): $t = 2.68$, $p < .02$.

Assessing the training effect must therefore take into account this difference in pre-training accuracy. To accomplish this, an improvement score was obtained for each teacher by subtracting pre-training per cent accurate from post-training per cent accurate. Comparing these per cent improvement scores, we find that the experimental training group had a mean gain in accuracy of +7.2 per cent, while the controls had a mean of -0.5 per cent. These means are significantly different: $t = 2.12$, $p < .05$.

It is obvious that the absolute gains obtained in the experimental

training group are relatively small. We are encouraged, however, for a number of reasons. First of all, the training technique used here can be unequivocally and objectively evaluated. This is seldom the case with standard teacher training programs. Secondly, the effect of the experimental training is demonstrated and significant. Thirdly, this statistically significant improvement was achieved with 6 to 8 hours of training, possibly the most encouraging feature of the study.

Thus we have significant evidence that accuracy in judging student comprehension on the basis of non-verbal cues can be improved by training in the use and interpretation of such cues. However, there are several ambiguities concerning the ability to make such judgments and the exact nature of the training effect. First of all, we must consider whether or not there is a pattern of non-verbal cues related to comprehension which is common across students. Our basic assumption has been that there is such a common set of cues, but until now we have had no data suitable for testing such an assumption. One of the interesting results of the earlier study referred to above was the lack of any evidence to suggest that teaching experience improves ability to use non-verbal cues to assess comprehension. However, teachers taking part in the earlier study had no opportunity to judge comprehension in clips of their own students. If there are very few, or no common cues of comprehension, they could not be expected to develop any generalized ability, even though they might become very adept at judging students whom they had taught. The previous study thus allowed no comparison of a general versus a student-specific ability to judge comprehension. The present data do allow for such a comparison since teachers' pre-training test

performance can be broken down by accuracy in judging their own students versus accuracy in judging students with whom they are unfamiliar. Table 12 (page 52) presents mean per cent accurate scores when they are broken down by own students versus others' students. The only significant effect is the one reflecting the fact, reported above, that controls had a higher pre-training mean than experimentals. There is no evidence that teachers judge their own students more accurately than they judge others' students. Per cent improvement scores were also broken down by own versus others' students, with similar results. The analysis of variance main effect of own versus others' students yielded $F < 1$. Thus there is no basis in either of these analyses to reject our assumption that the visual cues useful in judging student comprehension are common across students.

In describing the preparation of the post-training test films, above, it was pointed out that three types of clips were used in order to be able to check on two possible "practice effects" of the experimental training procedure. Actually, these effects can be thought of as alternative interpretations to a general training effect. They suggest task specific learning rather than improvement in general ability. The first practice effect suggested was that experimental teachers were simply learning the correct answers to the training clips because they were given right-wrong feedback during training. To paraphrase, the training procedure simply coached them for a retest. If this were true, we should find post-training performance on clips they were trained on superior to those on which they were not trained.

The second practice effect suggested is similar to the question

3.11 TABLE 12

Comparison of Experimental and Control Pre-Training Mean Per Cent Accurate Scores Broken Down by Own Versus Others' Students

	Own	Other	
E	55.6	53.0	E vs. C F = 4.94, p<.05 OW vs. OT F<1 Interaction F<1
C	60.0	59.7	

raised above concerning general versus student-specific cues of comprehension. This practice effect suggests that possibly experimental teachers are learning several sets of comprehension related cues, one for each student shown in the training sessions, again due to right-wrong feedback during training. If this were true, we would expect post-training performance to be better on all clips showing these students, regardless of whether the specific clips had been shown in training sessions. Table 13 (page 54) presents post-training per cent accurate scores when broken down by previous exposure to students shown in the film clips: Those clips shown in training; new clips on students shown in training, and new student clips not previously shown. It is evident that when we do not take into account pre-training differences in accuracy between experimentals and controls, no significant difference obtains in a post-training comparison. It is also evident that there is a "previous exposure" effect, suggesting both the effects discussed immediately above. Inspection of the means reveals, however, that the effect is equally present in experimental and control data, and thus cannot be invoked as an alternative to the interpretation that the experimental training procedure improves general ability to judge student comprehension.

To summarize, evidence has been obtained to support the hypothesis that the ability to judge student comprehension on the basis of non-verbal cues, while largely undeveloped in the normal classroom teacher, can be significantly improved by training in the recognition and interpretation of such non-verbal cues. Further, the obtained improvement in this ability cannot be dismissed as a transient, task-specific effect.

3.12 TABLE 13

Post-Training Per Cent Accurate Scores

Broken Down by Previous Exposure

	Training Clips on Training Students	New Clips on Training Students	New Clips on New Students
E	67.8	63.0	56.9
C	65.6	59.8	57.9

E vs. C $F<1$

Type of Clip $F = 4.10$, $p<.05$

Interaction $F<1$

3.13 VALIDITY OF NON-VERBAL CUES AS PREDICTORS OF STUDENT COMPREHENSION

The experimental training procedure described above and used in this study has been characterized as practice in "the correct scoring, interpretation and meaning of non-verbal cues," followed by feedback on the student's actual performance in each clip. Because the investigators themselves had no concrete basis for asserting definite relationships between the non-verbal cues and comprehension, a good deal of latitude existed in the "interpretation and meaning" judgments made by teachers during training. While we have found no evidence to suggest that the experimental training effect is a task-specific one, it is still possible that the only significant achievement of the training procedure was to force teachers to pay closer attention to student behavior in the clips. Without a doubt, there were individual clips where the relationship between cues and student performance was inconsistent with that suggested by the investigators. Thus it is possible that teachers were using different cues and/or interpretations of cues than suggested during training to judge comprehension.

It seemed important, therefore, to investigate the extent and nature of relationship between coded cues and both actual student performance and teacher judgments of student performance. Close inspection of the 22 specific cues included in the coding form revealed that five of these concerned characteristics of the students answering response. Since these cues are not available to teachers in an ordinary classroom situation, they were omitted from the analysis. Three

additional cues were dropped when data tabulation revealed no variability in the coded behavior.

Cue-coding data on 182 clips had been obtained during eight experimental training sessions with groups of five teachers. For each clip, the most frequently checked category on each cue was taken as a "group code" on that cue for that clip. This strategy was necessary due to the fact that although reasonable reliability was achieved, agreement was seldom perfect when cues were coded independently by teachers. The over-all ratio of agreement to total possible agreement among groups of five teachers coded individual items ranged from 63.5 per cent to 99.6 per cent with a mean of 81.1 per cent. Regression analyses were then performed to investigate relationships between cues and both judged and actual student performance.[6] Judged performance was also treated, in some of these analyses, as an independent variable predicting actual performance.

Column 3 of Table 14 (page 57) presents partial correlations between 15 independent predictors (14 cues and judged student performance) and actual student performance. The multiple r obtained is .40, and is significantly greater than zero by analysis of variance ($F = 2.12$, $p < .05$). Inspection of the individual partial r's reveals two noteworthy results. All partial r's are very small, the largest being between judged and actual performance. When judged performance is omitted as a predictor, the multiple r drops to .31, which is non-significant by analysis of variance.

---

6. These analyses were performed using the BIMD 06 program and the IBM 7090 computer at the Stanford Computation Center.

3.14 TABLE 14

Partial Correlations of 15 Predictors
of Student Performance With
Actual Student Performance

Description of Item	r's		
	Expect	Actual	Judged
1. Amount of time looked at source	+	.08	-.05
2. Number of times looked away from source	+	.02	.16
3. Speed of eye movements away from and returning to source (s, f, none)	+	-.06	.03
4. Amount of blinking	+	-.09	-.07
5. Grouping of blinking	?	.09	-.12
6. Duration of lowering eyebrows	+	-.05	.20
7. Strength of lowering eyebrows	+	.06	-.16
8. Duration of raising eyebrows	+	-.09	-.02
9. Strength of raising eyebrows	+	.03	.01
10. Movement of hands on face	+	-.15	.00
11. Frequency of general body movement	+	-.18	.14
12. Extent of general body movement	+	.17	-.15
13. Amount of mouth movement	+	.05	.04
14. Chewing gum or candy, y=1, n=2	+	.01	.05
15. Judged student comprehension R=2, W=1 Actual student comprehension R=2, W=1	+	.27	---

Secondly, two of the four largest r's are negative, where we had anticipated all positive r's. The partials remain largely unchanged in the second analysis where judged performance is omitted.

Several subsets of cues were selected and additional regression analyses were performed. The general effect of omitting various cues or sets of cues is to reduce the magnitude of the multiple r slightly. Only one subset analysis, using cues 1, 4, 5, 8, 10, 11, and 12 achieves a significant multiple r ($r = .28$, $p < .05$). The correlation of judged to actual performance is significant ($r = .26$, $F = 13.03$, $p < .001$), and inclusion of this item with any subset of cues generally improves both the multiple r and its significance.

The second set of regression analyses performed was to investigate the relationship between cues and judged answer. The multiple correlation of all cues as predictors of judged answer is significant but small ($r = .39$, $F = 2.15$, $p < .05$). Column 4 of Table 4 presents partial correlations of each cue with judged performance. Three different subsets of cues provide slightly smaller but highly significant multiple correlations. The best subset of predictors of judged performance were items 2, 5, 6, 7, 11, and 12, providing $r = .37$, $F = 4.52$, $p < .01$.

It is obvious that while some significant relationships exist between cues and judged performance, and between judged and actual performance, the predictors account for very little of the variance of the respective dependent variable. The most reasonable interpretation of these results is that there is very little evidence that the cues used in this study are related in any significant way to actual student performance. There is, on the other hand, more substantial evidence that teachers were

basing their judgments of student performance on some of these cues.

One could speculate endlessly on the proper interpretation of the obtained correlations; particularly with respect to the signs of the partial r's and sign changes between columns 3 and 4 of Table 14. However, in light of the size of these correlations, such speculations are probably best held to a minimum. It is worthwhile to point out that two sources of substantial unreliability exist in these data. First of all, there is an unknown amount of unreliability in the students' actual responses. Secondly there is probably even greater unreliability in the teachers' cue-coding data. Nothing short of improving the reliability of the student comprehension measure and reshooting the films can be done to reduce the first source of extraneous variability. On the other hand, a more careful recoding of the non-verbal cues in the present film attacks the larger source of extraneous variability.

The basic problem of ascertaining just which patterns of students' non-verbal behavior indicates comprehension or its lack needs further attention. Our research to date indicates that (1) adequate examples of such students' behavior can be obtained, (2) film protocols of such behavior can be reliably coded, and (3) teachers trained in coding such protocols improve somewhat their ability to make correct judgments of student comprehension. However, it is also clear from our efforts to date, that although the student behavior that underlies teachers' judgments of comprehension can be identified, we have not yet succeeded in isolating which student behaviors indicate actual degree of comprehension.

Reliability of teacher ratings of student behavior recorded in clips is one thing, but the extent to which the behavior is being

accurately coded may be quite another. Teachers may agree, for example, that little or no eye-blinking occurs because no one of them noticed such behavior. The next step then is to obtain a set of coded film protocols in which replay of each clip has occurred sufficiently to guarantee that the actual student behavior is being accurately coded. It should then be possible to discover the relative contribution of each of the coded student behaviors to actual student understanding.

IV. PROBLEMS OF SOUND FILM RECORDINGS IN HIGH SCHOOL CLASSROOMS

One of the primary objectives of this project was to develop the ability to make working quality sound motion pictures in high school classrooms without unduly disturbing the procedure of the class and especially without making students' behavior highly artificial. There are a number of problems involved in doing this. They can best be treated under the two components of the film making process, i.e., picture and sound.

4.1 RECORDING THE VISUAL IMAGE

An early test of visual coverage was done at the Carlemont High School in San Carlos, California. At that time it was felt that no interference at all with the normal activities in the calssroom would be the optimal conditions for obtaining visual data. For this test a ground level classroom was chosen and the camera was mounted outside the classroom looking in through a window. The camera was fitted with a 15mm lens, the angle of which was wide enough to include most of the students in the picture and also a large part of the activity of the teacher. It was obvious from the results of this test that the camera gave us a record of what was going on in the classroom with minimal disruption or distraction but (1) the rigid distance from the camera to the students and (2) the fact that the camera could not be moved (nor could the apparent distance from the camera to the subject be varied) made the films of negligible value. It was at this point that we decided to place both the camera and the camera operator inside the classroom.

It was our intention to use two cameras in the room, one in the front, the other in the rear such that the rear camera would continually keep a record of the teacher and what the teacher was doing and the front camera would keep a continual record of the class. To this end, two Auricon Cine-Voice cameras were specially modified. The modification consisted essentially (1) of strengthening the structural members of the camera such that it was reliable for runs up to 33 minutes without a break and (2) adapting the cameras so that they would accept 1200 foot magazines rather than the 200 or 400 foot rolls for which they were originally designed.

In addition, a problem was encountered in accommodating the mass of film that accumulated on the takeup side of the magazine and keeping it under a constant rate of tension even though the speed of the takeup mechanism varied. W. A. Palmer Films, Inc., in San Francisco, was of great value in engineering this modification. They suggested adapting to the camera a separate torque motor powered by a varialbe auto-transformer which would operate the magazine takeup mechanism. The autotransformer, in turn, was controlled by a feedback mechanism consisting of an arm riding on the film taken up in the magazine so that as the amount of film in the takeup side of the magazine increased the arm actuated the autotransformer which in turu increased the voltage to the torque motor, so that a constant torque was applied to the hub of the takeup spindle. This proved to be of extraordinary value and was quite reliable once an oil immersion-type capacitor was substituted for the originally used electrolytic capacitor in the autotransformer torque motor drive unit.

One other modification that was done on one of the cameras was to fit it with a viscous dashpot so that the variable focal length lens could be driven smoothly. The other camera was not so fitted because it was felt that in the photography of the teacher there would be little need for very much shift in apparent distance between camera and subject. After a number of tests, the Angenieux 17-70mm variable focal length lens was chosen. These lenses proved to be simplest to operate and also considerably sharper than other lenses of this kind tested.

Perhaps the major difficulty with lenses of variable focal length is that they tend to be slightly out of focus during the time that the focal length of the lens is being changed (that is to say, during the "zoom"). Also, although the lens may be focused correctly at one particular focal length, the question of whether this focus will hold as the lens is manipulated through a variety of focal lengths is also one that has not been solved satisfactorily by the majority of lens manufacturers. We feel that the Angenieux lens gave us optimal results.

In order to register an image on film it is obviously necessary for there to be light in the classroom. This proved to be an important problem for us. The emulsion chosen for our work was Eastman Type 7276, Tri-X Reversal 16mm motion picture film. Previous to deciding on Type 7276 we had experimented with a number of different emulsions both negative and reversal. Our concern was to choose an emulsion which would give us the maximum light sensitivity compatible with high information storage characteristics. That is to say, we were interested in a film which would operate under extremely low light conditions and still give us a very sharp and well-resolved image. In our tests we used Eastman

Type 7231, Plus-X Negative film and Eastman Type 7222 Double X Negative film as well as Gevaert High Speed Panchromatic Reversal Film, Type 880. We found it possible to use the Type 7278 Eastman Tri-X Reversal film with a minimum ambient light level of 10 footcandles as compared to the minimum needed ambient light levels of Eastman 7231 which was 40 footcandles and Eastman Type 7222 which was 13 footcandles. In addition, Gavaert Type 880 required approximately 18 footcandles to come up with a comparable exposure.

Because of the great light sensitivity of the film chosen it was possible for us to work under any classroom lighting conditions with no additional lights other than those present in the fixtures extant in the classroom. In order to relieve operator fatigue and to give pictures of maximum stability we used hydraulically damped tripod heads. The cameras fitted with 1200 foot magazines and outboard torque motors for the magazines and loaded with 1200 feet of film weighed in excess of 45 pounds and it was therefore necessary to use relatively heavy-duty tripod and pan head equipment. We found the most satisfactory combination to be the Mitchell heavy-duty tripod and the O'Connor Model 05 tripod head for the camera in the front of the room - this was the camera used to cover the students - which was subject to the greatest variety of movement. The camera in the rear of the room to record teacher behavior was fitted on a heavy-duty Miller fluid head which was in turn mounted on Pro-Junior tripod legs. It was our experience that this was at best a makeshift arrangement, but it was used in order to keep costs down and portability at a maximum.

Once the image was registered on the camera original film it was

standard procedure to deliver the film, exposed but unprocessed, to W. A. Palmer Films, Inc., in San Francisco which did the processing. The problems in processing Eastman Type 7278 Tri-X Reversal film, or any reversal film for that matter, are rather extraordinary in that scrupulous control over the quality of the processing solutions must be maintained at all times. There is virtually no allowable variability. In 7278 any variability in the specified strengths of processing solutions may be manifested as a clumping together of the silver grains on the film which renders the image very grainy and reduces the resolving power of the film's emulsion. Using a constant replenishment system and keeping the processing temperature at a constant 75 degrees, the Palmer Laboratories processed the film for us so that it maintained a resolving ability of approximately 100 lines per millimeter. This fidelity enabled us to examine relatively small portions of the frame with considerable accuracy. The original was then work-printed on Eastman Type 7371 Reversal Duplicating film and suffered only minor loss of quality, resulting from the unavoidable increase in image contrast due to the printing process

4.2 RECORDING THE SOUND

The problems of sound recordings under classroom conditions has proved to be the most difficult ones for us to solve. The acoustic treatment of high school classrooms varies enormously depending on the age and construction of the school building. Uniform, however, is the fact that classrooms normally have hard floors, desk tops and rather large areas of blackboard which are extremely reflective surfaces and cause sound to be distorted in a number of ways.

The problem of microphoning the teacher was solved through the use of a Vega wireless microphone. This microphone is approximately three inches long, an inch and a quarter in diameter and contains both a microphone element and a miniaturized FM transmitter so that the microphone need not be connected directly by wire to the recorder. This gives the teacher as much flexibility as he would have without any microphone on at all. The signal from the miniaturized FM transmitter is received by a highly sensitive and discriminative FM receiver which in turn is fed into a mixer and thence into a recorder.

We tried a number of different ways to secure adequate sound from the students before finally eliminating some of the more troublesome problems. At first it was thought that an ultradirectional microphone could be used in conjunction with the camera in the front of the room. We tested ultradirectional microphones made by several different manufacturers, including Electro-Voice and ESL, and found that they were inadequate for our purposes whether mounted on floor booms or on top of the cameras themselves. The ratio of sound picked up from the front of the microphone to that from the back of the microphone was such that if the mike were mounted directly on the magazine the noise would be objectionable. It should be noted that the Auricon cameras run, for all practical purposes, silently. That is to say that one may approach the camera to a distance of about three feet without hearing the noise of the operation. This is a tolerable limit. Mounting the microphone on top of the camera, however, imposes noise which is not tolerable. Further, it was found that these microphones were not particularly sensitive and a certain amount of amplification noise was present when the

sound level of the microphones was increased.

The alternative to mounting the microphone on top of the camera was to place it somewhere else in the front of the room. This introduced two more difficulties: first, the synchronization of the movement of the microphone with the movement of the camera and second, the introduction of yet a third technician into the classroom.

Another alternative that was tried was the development of a parabolic microphone which would use a parabolic reflector and a relatively inexpensive microphone which would be used to pick up the sound concentrated by the reflector. This was abandoned when it became evident that the differential distances from the microphone to the various students in the classrooms was such that an undue amount of adjustment would be necessary in order to operate the parabolic microphone effectively.

Still another alternative was attempted. A number of high output crystal microphones were spotted around the classroom. This plan proved to be a relatively unsatisfactory one because of the amount of hiss and extraneous noise produced by the microphones and also because of the fact that the microphones tended to pick up unwanted sounds such as students tapping their desks, books shuffling and shoes scraping the floor. Had the crystal microphones been able to give us very high quality sound from the students' voices, then the extraneous noise would have been tolerable, but the inherent design capabilities of crystal microphones are such that this quality is not possible.

The ultimate solution was to use four extremely high quality microphones, omnidirectional in nature. We experimented with a number of microphones, and decided finally on the Hartsfield microphone which was

only recently developed by a technician of the National Bureau of Standards and which is custom made by the inventor for special sound problems of the kind we faced. The extremely small size of the diaphragm in this microphone and its extraordinarily high degree of compliance makes it possible for the microphone to record accurately extremely complex sound patterns with almost no measurable distortion. Further, these microphones were equipped with rolloff filters so that sounds below 200 cycles, well below the predominant frequencies of the human voice, could be excluded. This minimized a great amount of extraneous noise.

Ultimately the configuration of microphones in the classroom came to be as follows: the teacher's microphone is a Vega wireless which is fed into the mixer and then into the recorder; spaced around the classroom on stands are four Hartsfield microphones. The microphones are all fed into a selector box. At any one time only one of these microphones can be actuated by a button on the selector box, but extremely rapid changes from one microphone to another can be accomplished. In this way it is possible for us selectively to microphone 1/4 of the classroom at a time, thus reducing the amount of extraneous noise. From the selector box the chosen microphone signal is then sent to a mixer and then into the recorder. The sound is recorded on a Nagra Type 3 quarter inch synchronous pulse recorder operating at a speed of 7 1/2 inches per second. One of the features of this recorder is an automatic limiter, so that the sound level in the classroom is made fairly constant for practical purposes by the limiting action of the recorder. The sound is transferred from the 1/4 inch tape to 16mm magnetic film.

The 16mm magnetic film is then sent to the laboratory together with the camera original, and the laboratory then makes a composite print with an electro-printed sound track.

This equipment gives us a combination of high portability and extremely high quality. It is possible to set up in a classroom in less than 10 minutes if the microphones' positions and the power cables for the cameras are taken care of after school the previous day or before school on the day of shooting. This pre-setup enables us to get into classrooms during the normal break between classes. The nature of the equipment enables us to do this work using a crew of three; one man on each camera and one man on the sound. Such wiring as may be necessary, including microphone cord, wiring cables and accessories, are run around the periphery of the room so that there is no traffic problem as far as the students entering or leaving the room is concerned. The teacher, with his wireless microphone, is free to move anywhere at will. The cameras are placed in the room and operators present for as much as three consecutive classroom days during the same class period. Because of the quiet cameras students cannot tell when the cameras are in operation. Thus although the students know that some of the time they will be filmed, they don't know exactly when the cameras are operating. This familiarization process plus the fact that the students do not know exactly when they are being filmed minimizes but does not entirely eliminate the effect of the recording process on student behavior. These factors make it possible for us to produce high quality films of classroom teaching with a minimum amount of camera consciousness on the part of the students and the teachers.

4.3 SOUND FILM RECORDING IN LABORATORY SITUATIONS

For intensive analysis and the generating of experimental materials, a considerable amount of shooting was done in laboratory situations, so that stimuli could be controlled, and the relationship between camera and subject could be adjusted without regard to the exigencies of an on-going class. Lighting could be more carefully controlled, and could be used to emphasize various kinds of visual cues.

The laboratory shooting was done in a large radio studio on the Stanford campus which, because of its excellent acoustics, was the most suitable room available. Ten students were, over a period of several weeks, seated in front of a blackboard and a camera, and were taught material in mathematics which was beyond the level they had attained in the course of their junior high school curriculum. The camera was visible to the students at all times.

The cameras used were the specially adapted Auricons previously referred to, equipped with Angenieux 17-68mm variable focal length lenses. During most of this work, the lens was extended close to its maximum focal length, so that a full face close-up of each student filled the screen. The teacher was out of camera range.

Several lighting styles were used:

FLAT LIGHTING which utilized a #2 Photoflood in a Smith-Victor 16' reflector and an aluminized sheet reflector gave us a reflectance ratio of approximately 2:3 between one side of the subject's face and the other.

LOW KEY LIGHTING was used to accentuate the facial muscles of some subjects, employing various arrangements of photo flood and reflector flood

and spot units. Subsequently Color-tran units were required in order to vary the contrast ratio and intensity without moving the light units. BACK LIGHTING was used to accentuate gross postural changes in some subjects.

Generally it was found that the most satisfactory lighting results were obtained by the use of a relatively flat lighting treatment, employing a kicker light when the top of the subject's head was in danger of blending with the background. This procedure enabled us to make more accurate judgments in the extent of minor head and body movements. A moderate shadow-to-key light contrast ratio afforded sufficient gradation to discern variations in facial expression without washing out key shadow producing areas around the mouth, the eyes and the forehead. For the purpose of analyzing the movement in facial muscles, a direct head-on view (so that both eyes are visible) is desirable, but since this kind of arrangement was not feasible, a compromise between a full face and profile view resulted in satisfactory though not optimal coverage for our purposes. Flat lighting poses a problem of causing the subject to squint or blink with an unusually high frequency, thus the level of the illumination must be kept to the minimum acceptable for exposure, and the lighting units placed as far off center as possible without resulting in drop-off.

The film was processed as described in the previous section.

Sound was picked up from the subject and the experimenter (using an ElectroVoice model 664 microphone) and fed directly into a MagnaSync type 5 16mm film recorder.

A number of other laboratory films were shot using two cameras and two subjects simultaneously, so that differential reactions to a common input could be compared. The procedure was essentially as outlined above, except for the fact that sound was picked up from the experimenter only, using the Vega Wireless Microphone, and the cameras were placed immediately adjacent to each other and to the side of the blackboard on which the instructor was writing.

FIGURE 3

4.4 Laboratory Filming Procedure

V. SUMMARY

In order to investigate to what extent sound film recordings of pupils in teaching-learning situations could be used as a basis for training teachers to improve their accuracy in interpreting non-verbal communications from students, equipment and methods were devised to record instances of student classroom response and several experiments were performed.

The first phase of the research was concerned with the investigation of the relative value of verbal and non-verbal cues as they contribute to the accuracy with which teachers make judgments concerning student comprehension of the material being presented. Non-teachers, inexperienced teachers, and experienced teachers were shown sound film clips of student-teacher interactions and were asked to rate the amount of student comprehension evident in the clips. Some subjects were shown only the picture track, some subjects heard only the sound track, and some subjects saw and heard the complete sound-film recording. The results strongly indicate that verbal cues contribute significantly to the accuracy of judgment, while the presence of non-verbal cues, specifically visual ones, do not increase accuracy of judgment above chance expectancy. Also, no evidence was found to support the hypothesis that normal teacher training or teaching experience making content-identical judgments improve the accuracy of those judgments. It was shown that low accuracy was accompanied by low confidence in the judgments made.

In the second phase of the study, the experimenters set out to investigate whether accuracy in judgment of student response by teachers

could be improved by training. A number of clips of student response to a controlled lesson were generated, and subjects were pre-tested on ability to judge student response from viewing silent clips. Pre-test results indicated accuracy consistent with earlier findings, not significantly different from chance. Pre-test group randomly split into training and control group. The training group was given 6-8 hours of training on a 24 item coding scale devised by the experimenters, which dealt with such items as orientation toward source, frequency of movement, facial gestures, etc. Control group received placebo training by viewing films relating to interpersonal communication and discussing those aspects of the film which were relevant to the student-teacher communication process. Both groups were given a post-training test which indicated that training significantly improved accuracy of teachers, while $\underline{S}$'s in placebo condition showed no difference. There is significant evidence that accuracy in judging student comprehension on the basis of non-verbal cues can be improved. Further, there is no evidence that teachers judge their own students more accurately than they judge others' students, supporting the assumption that visual cues of comprehension are common across students.

The general area of student non-verbal comprehension patterns needs further research, although there is evidence to indicate that (1) adequate examples of such behavior can be obtained; (2) film protocols can be coded reliably; (3) teachers trained in coding improve in their judgmental ability of student comprehension. Investigation continues relative to the isolation of which behaviors indicate actual degree of comprehension. Further analysis of relative contributions of

behaviors should be examined with the clear understanding that reliability and validity are quite distinct, and usual coding reliability measures are not completely applicable to the analysis of moving visual phenomena.

VI. PERSONNEL

Principal Investigator - Nathan Maccoby, Ph.D., Professor of Communication and Senior Staff Member, Institute for Communication Research, Stanford University, Stanford, California.

Associate Investigator - Henry Breitrose, M.A., Instructor in Communication and Research Associate, Institute for Communication Research, Stanford University, Stanford, California.

Research Associates: Jon Jecker, Ph.D., Assistant Professor, Department of Psychology, Kent State University, Kent, Ohio.

Ernest D. Rose, M.A., Visual Communication Specialist, University of California Extension Division, Berkeley, California.

Peter Gumpert, M.A., now in the Department of Psychology, University of North Carolina, Chapel Hill, North Carolina.

Research Assistants: Richard T. Gilbert, M.A.
Jerry Galyean, M.A.
David Geeting, M.A.

Data Processing Specialist - Lee Ruggels, M.A., Research Assistant, Institute for Communication Research, Stanford University.

Secretaries: Gracie Barron, Linda Miller, Jean Bull, Garland Robinson.

Consultants: Arthur Lumsdaine, Ph.D., Professor, School of Education, University of California at Los Angeles.

Dwight Allen, Ed.D., Assistant Professor, School of Education, Stanford University, Stanford, California.

A. Kimball Romney, Associate Professor, Department of Anthropology, Stanford University, Stanford, California.